Proc. Natl. Acad. Sci. USA
Vol. 75, No. 1, pp. 136–139, January 1978
Biochemistry

Role of the host in virus assembly: Cloning of the *Escherichia coli groE* gene and identification of its protein product

(transducing phage/*in vitro* recombination)

ROGER W. HENDRIX AND LAPCHEE TSUI

Department of Life Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

Communicated by Armin Dale Kaiser, October 14, 1977

ABSTRACT **Correct assembly of the heads of bacteriophages λ and T4 requires the function of the *groE* gene of the *Escherichia coli* host. We have isolated a transducing derivative of λ, called λgt-Ec-groE, that carries a functional copy of the *groE* gene. Unlike wild-type λ, this phage is able to form plaques on hosts with a mutant *groE* gene. We have isolated an amber mutation in the *groE* gene carried by the phage, and this has made it possible to identify the *groE* product as a protein of molecular weight 65,000. In the phage, the *groE* gene is under the control of an early phage promoter.**

The assembly of structurally complex viruses such as the large bacteriophages is more than a simple self-assembly process. Although much of the information for assembly is carried in the structure of the molecules being assembled, correct assembly also typically requires participation by molecules that are not components of the assembled virion. There are now numerous examples of phage-coded proteins that participate obligatorily but transiently in virion assembly, bringing about both covalent and non-covalent alterations of the intermediates of the assembly pathway. In addition, it is clear that components of the host cell are also involved directly in the assembly of phage-coded molecules into virions. (For recent reviews of these aspects of virus assembly, see refs. 1 and 2.)

The most extensively studied case of a bacterial gene involved in virus assembly is the *groE* gene (also called *tabB* or *mop*) of *Escherichia coli*. A functional *groE* gene is required for correct assembly of heads of both λ and T4. (3–8). If cells that carry a mutant *groE* gene are infected with wild-type phage, the head proteins are assembled incorrectly, and the processing of head proteins that normally occurs fails to take place. All other aspects of the phage growth cycles are normal, arguing that the product of the *groE* gene must participate directly in the head assembly process. Phage mutants that overcome the *groE* block, called ϵ mutants, have been isolated and mapped. In the case of λ, ϵ mutants map in either gene *E* or gene *B* of the phage. These genetic data have led to the suggestion that the *groE* gene product interacts with the gene *E* protein (gpE) and the gene *B* protein (gpB), both of which are components of the λ head. ϵ mutations in T4 map either in gene 23, which codes for the major head subunit, or in gene *31*, which is required for head assembly but whose protein product is not a part of the finished phage. When T4 head assembly is blocked at the level of gene *31* action, by mutation of either *groE* or gene *31*, head proteins are found associated with the cell membrane. This property has raised the possibility that the *groE* gene product might be a membrane component or might interact with the membrane.

Studies aimed at elucidating the detailed biochemical mechanisms by which *groE* participates in phage assembly have been severely hampered by the fact that the *groE* gene product has not been identified. In this paper we describe experiments that have allowed us to clone the *groE* gene and to identify the protein for which it codes.

MATERIALS AND METHODS

Phage and Bacterial Strains. λgt-Ec-groE was selected as described in *Results*. This phage carries the cIts857 and *nin*5 mutations. λgt-Ec-groE cI^+ and λgt-Ec-groE *Q*am21 were constructed by recombination with λ^+ and λ *Q*am21 *c*Its857. λgt-Ec-groE cI^+ has lost the *nin*5 deletion as judged by the sensitivity of the virion to inactivation by pyrophosphate; λgt-Ec-groE *Q*am21 was not tested for *nin*5. The *E. coli groE* strains used have been described previously (3).

Phage Protein Labeling. Phage-coded proteins were labeled by the method described by Hendrix (9) with the following modifications. Cells were grown in RG maltose medium (3) containing 0.1 mM Na_2SO_4. The radioactive label was $H_2{}^{35}SO_4$ from New England Nuclear. Each sample received 250 or 500 μCi of label. Cells were harvested 1 hr after infection by pelleting and resuspending in 200 μl of gel sample buffer. For the pulse-labeling experiment, aliquots of an infected culture received 250 μCi of label at 0, 10, 20, 30, and 40 min after infection. Each aliquot was harvested as above 10 min after addition of label.

Selection of λgt-Ec-groE Amber Mutants. *E. coli* strain 594 was grown to 2×10^8 cells per ml in M9 medium supplemented with 0.4% maltose and infected with λgt-Ec-groE cI^+ at a multiplicity of infection of 4. Adsorption was carried out at 0° for 15 min in the presence of 20 mM $MgCl_2$. The mixture was then diluted 10-fold in M9 maltose medium, and *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine was added to a final concentration of 5 μg/ml. The infected culture was shaken at 37° for 90 min and chloroform was added to ensure complete lysis. This level of mutagenesis produced about 10% clear plaque mutants in the progeny. The lysate was plated on strain 594 and turbid plaques were transferred by toothpick to plates seeded with appropriate indicator bacteria. The plates were incubated overnight at 30°. The phage that gave a lysis zone on 594 and *groE*A16 with the *supE* suppressor, but not on *groE*A44, were scraped from the plate, resuspended in λ dilution buffer, and plated on the three indicating bacteria. Two out of 1830 plaques screened had the *groE* amber phenotype. Six single plaques were isolated from each of the two candidates and plated again on the indicating bacterial strains. All of them showed the *groE* amber phenotype. The original λgt-Ec-groE cI^+ phage, as a control, plated equally well on the three hosts. Two single-

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: gpE, gene product of the λ *E* gene; gp groE, gene product of the *E. coli groE* gene.

Table 1. Plaque formation on *groE* hosts

Host	Phage				
	λε+	λεA36	λεA30	λεB764	λgt-Ec-groE
gro+	+	+	+	+	+
*groE*A36	−	+	+	−	+
*groE*A44	−	−	+	−	+
*groE*B764	−	−	−	+	+

Plating properties of λgt-Ec-groE and λε mutants. A + indicates a plating efficiency equal to the plating efficiency on a *groE*+ strain. A − indicates a plating efficiency at least 10^5 times lower than on a *gro*+ strain. In addition to the strains shown, λgt-Ec-groE was tested on and found to grow on the following strains: *groE*A30, *groE*114, *groE*A140.

plaque isolates, *am*11 and *am*21, one from each original isolate, were used for further experiments.

Phage Density Measurements. λgt-Ec-groE *c*I+, λ*b*2 *c*I26, and λ*imm*434 Sam7 were mixed, and the solution was brought to a density of approximately 1.5 g/cm^3 by adding a saturated solution of CsCl. The mixture was centrifuged at 24,000 rpm for 36 hr in a Beckman SW 60 rotor, and fractions were collected and titered on bacterial strains *groE*A140, Ymel (λ*imm*434), and Ymel (λ). The λ*imm*434 Sam7 and λ*b*2 *c*I26 markers were separated by 9.4 fractions, and λgt-Ec-groE *c*I+ banded 2.3 fractions lighter than λ*imm*434 Sam7. λ*imm*434 Sam7 was taken to have a 2.0% deletion relative to λ+, λ*b*2 *c*I26 was taken to have a 12.5% deletion, and λgt *nin*+ was taken to have a 21.1% deletion (10, 11).

RESULTS

Isolation and characterization of a *groE* transducing phage

Wild-type λ fails to form plaques on *groE* strains of *E. coli* because of the mutation in the *groE* gene of the host. A phage that carried the wild-type allele of *groE* in its chromosome might be expected to overcome the effects of a defective *groE* gene in the host and form plaques on a *groE* strain. We were led to test this hypothesis by the availability of a pool of λ transducing phages carrying various different segments of the *E. coli* chromosome. The phage pool is the one described by Cameron *et al.* (12). It was made by digesting *E. coli* DNA with *Eco*RI restriction endonuclease and inserting the resulting fragments into the λgt vector phage. This pool of phages is the one from which Cameron *et al.* isolated a DNA ligase transducing phage, and it should in theory contain transducing phages representing all possible *Eco*RI fragments of *E. coli* DNA, with the exception of fragments that confer a growth disadvantage on the phage carrying them.

The phage pool was plated on *E. coli groE*B515, and plaques appeared at a frequency of approximately one per 10^4 phage plated. The plaques were roughly the size of wild-type plaques and were uniform. Stocks were grown from five of the plaques, and, when preliminary experiments showed no differences between them, one stock was used for all subsequent work. Following the nomenclature of Thomas *et al.* (10), the phage obtained was named λgt-Ec-groE. This name indicates the vector phage (λgt), the source of cloned DNA (Ec = *E. coli*), and the method of selection (growth on a *groE* strain).

Table 1 shows the plating properties of λgt-Ec-groE and three typical λε mutants. The ε mutants, which have been described previously (3), contain a mutation in λ gene *E* or *B* that enables the phage to plate on certain *groE* hosts. λgt-Ec-groE differs from λε mutants in several respects. As Table 1 shows, λgt-Ec-groE plates on all *groE* strains tested, whereas all known λεs plate on only a subset of *groE* strains. Furthermore λεs frequently are able to make plaques on *groE* only at 30°. In contrast, we find that λgt-Ec-groE forms plaques on *groE* strains at 30°, 37°, and, provided that the *groE* strain is not temperature sensitive, at 42°. In addition to these differences in plating behavior, the frequency at which λgt-Ec-groE arose is 3–5 orders of magnitude higher than the spontaneous frequencies reported for ε mutations. From these properties of λgt-Ec-groE, we conclude that it is not a λε, and therefore that its ability to form plaques on a *groE* strain is conferred by the inserted *E. coli* DNA.

To determine whether the inserted DNA of λgt-Ec-groE actually carries the *groE* gene, we tested whether λgt-Ec-groE could transduce a *groE*− strain to *gro*+. *E. coli groE*A44 fails to support growth of both λ and T4 and is itself temperature sensitive with respect to growth. All three of these properties are consequences of the mutation in its *groE* gene (13). We infected *groE*A44 with a *c*I+ derivative of λgt-Ec-groE, and selected colonies that could grow at 42°. Temperature-resistant colonies appeared at a frequency of 3×10^{-4}, which is 10^4-fold above the reversion rate of *groE*A44. Of 36 such colonies tested, all had been transduced to *gro*+ by the following criteria: they were temperature resistant (by selection), they had λ immunity, and they supported growth of T4 with normal efficiency. We conclude that λgt-Ec-groE carries the *E. coli groE* gene in its DNA.

The size of the inserted DNA in λgt-Ec-groE was estimated by measuring the density of the λgt-Ec-groE virion. It was banded in a CsCl equilibrium density gradient along with density markers λ*imm*434 and λ*b*2. λgt-Ec-groE banded at a position corresponding to a net deletion of 4.6% relative to wild type λ. From the known size of the λgt vector DNA (14), we calculate that the inserted DNA is equivalent to 16.5 ± 0.5% of wild-type λ DNA, or 7.9 ± 0.2 kilobases. This is sufficient DNA to code for about 300,000 daltons of protein.

Proteins made by the *groE* transducing phage

Proteins encoded by λ phages can be labeled radioactively in cells that have been irradiated with ultraviolet light to reduce host synthesis after infection (9). Fig. 1 shows the results of such an experiment. Irradiated cells infected with λgt-Ec-groE or λ*b*2, or uninfected, were labeled with $^{35}SO_4$ from 0 to 60 min following infection, then solubilized and electrophoresed in a sodium dodecyl sulfate/polyacrylamide gel. [The ideal phage for the control infection would be λgt-0, the λgt vector with no inserted DNA, and not λ*b*2; however, λgt-0 does not exist as an infectious virion because the DNA is too small to be packaged (10).] Fig. 1A shows, first, that λgt-Ec-groE and λ*b*2 share a number of phage-specific bands. In addition, λ*b*2 makes three proteins that are absent in the λgt-Ec-groE lysate and λgt-Ec-groE makes one of about 65,000 daltons that is absent in the λ*b*2 lysate. The proteins that are specific to λ*b*2 are coded by a region of the λ DNA that is present in λ*b*2 and deleted in λgt (14, 9). The λgt-Ec-groE-specific band is not one that is normally made by λ, and must therefore come from the inserted *E. coli* DNA of λgt-Ec-groE. Fig. 1B shows the same experiment performed with cells that received a lower dose of irradiation. In this case, a substantial amount of cellular protein is made, including a protein that comigrates with the 65,000-dalton protein coded by λgt-Ec-groE. This suggests that the 65,000-dalton protein may be the same as a relatively abundant component of uninfected cells. However, the question of whether these two proteins are in fact identical will require further testing. The 65,000-dalton band shown in Fig. 1 is invariably made in a substantial amount in infections of λ-sen-

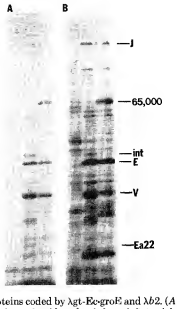

FIG. 1. Proteins coded by λgt-Ec-groE and λb2. (*A*) Sodium dodecyl sulfate/polyacrylamide gels of, from left to right, uninfected cells, λb2-infected cells, and λgt-Ec-groE-infected cells. The cells received 12,000 erg/mm^2 of ultraviolet irradiation prior to infection. (*B*) as in *A*, but the dose of irradiation was 2000 erg/mm^2. Bands that are identified in the figures but not in the text correspond to previously identified phage-coded proteins (9, 15).

sitive cells by λgt-Ec-groE. In most such experiments we also see a second λgt-Ec-groE-specific band at about 75,000 daltons (see Figs. 2–4). In a few cases, we have seen a third band at about 63,000 daltons, but its appearance has been too irreproducible to allow us to characterize it.

If λgt-Ec-groE is used to infect an irradiated cell that has λ immunity, the λgt-Ec-groE-specific bands are not seen above the level present in uninfected cells (data not shown). This argues that they are under the control of one or more of the λ promoters. If they also carry their own promoters in

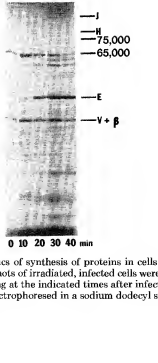

FIG. 2. Kinetics of synthesis of proteins in cells infected with λgt-Ec-groE. Aliquots of irradiated, infected cells were pulse labeled for 10 min, starting at the indicated times after infection, and then harvested and electrophoresed in a sodium dodecyl sulfate/polyacrylamide gel.

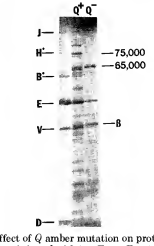

FIG. 3. Effect of *Q* amber mutation on protein synthesis. Irradiated cells were infected with λgt-Ec-groE or λgt-Ec-groE *Q*am21. Labeled virions are included on the left to provide molecular weight markers.

λgt-Ec-groE, these must be much weaker than the λ promoters.

The kinetics of synthesis of the 65,000- and 75,000-dalton proteins are shown in Fig. 2. Aliquots of irradiated cells infected with λgt-Ec-groE were pulse labeled for successive 10-min periods and then electrophoresed. The 65,000-dalton protein is synthesized with kinetics characteristic of early λ proteins, while the 75,000-dalton protein is synthesized with typical late kinetics. We conclude that the 65,000-dalton protein is under the control of an early λ promoter. Given the structure of the λgt vector DNA, this must be the P_L promoter. The 75,000-dalton protein appears to be under the control of the λ late promoter.

These results are corroborated by the experiment shown in Fig. 3, which compares the gel pattern obtained from λgt-Ec-groE to that from a λgt-Ec-groE derivative that carries an amber mutation in the λ *Q* gene. The *Q* gene codes for a positive regulator of λ late transcription. The introduction of a *Q* mutation into λgt-Ec-groE causes a marked reduction in the 75,000-dalton band but no significant change in the 65,000-dalton band. This again argues that the 65,000-dalton protein is controlled as an early λ protein and the 75,000-dalton protein as a late λ protein.

Identification of the *groE* protein

In order to identify the *groE* protein, we sought an amber mutation in the *groE* gene carried by λgt-Ec-groE. λgt-Ec-groE was mutagenized with nitrosoguanidine, and the surviving phages were screened for mutants that required an amber suppressor for growth in a *groE*$^-$ strain but not in a *gro*$^+$ strain. Of 1830 plaques screened, two showed the properties expected for a λgt-Ec-groE amber. In all subsequent tests both isolates behaved identically, and they may well be siblings.

Fig. 4 shows gels of irradiated cells infected with the two mutants, λgt-Ec-groE am11 and λgt-Ec-groE am21. In the amber phages both the 65,000-dalton and 75,000-dalton bands are missing. However, when the experiment is carried out in cells carrying the *supD* or *supE* amber suppressors, the 65,000-dalton band returns, while the 75,000-dalton band remains absent. The correlation between suppressibility of the phage phenotype and suppressibility of the 65,000-dalton protein argues strongly that the 65,000-dalton protein is the product of the *groE* gene.

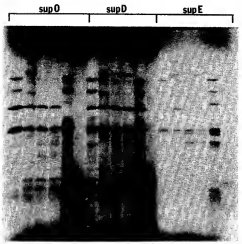

FIG. 4. Effects of *groE* amber mutation in strains with and without suppressors. The six gel columns in the first group are lysates from a *sup0* nonsuppressor strain. The second group of six is from a *supD* amber suppressor strain, and the third group of six is from a *supE* amber suppressor strain. Each strain was infected as follows, from left to right: λSam7 cIts857, λgt-Ec-groE cI+, λgt-Ec-groE am11, λgt-Ec-groE am21, λEam4 Sam7 cIts857, uninfected.

DISCUSSION

Probably the most important result reported here is the identification of the 65,000-dalton protein as the product of the *groE* gene. This identification rests on the following facts. It is one of the two proteins that can be identified as coded by λgt-Ec-groE but not by λ wild type, and it is therefore assigned to the piece of *E. coli* DNA inserted into λgt-Ec-groE. If the *groE* gene, which is on the inserted DNA, carries an amber mutation, the 65,000-dalton protein is not made in a nonsuppressing *sup0* strain. If the *groE* amber phage infects an amber suppressing strain, synthesis of the 65,000-dalton protein is restored to roughly half of the wild-type level. This identification is strengthened by results obtained by Georgopoulos and Hohn (16). They have isolated a similar *groE* transducing phage and have isolated missense mutations in the *groE* gene on that phage. They find that the mobility of the 65,000-dalton protein on sodium dodecyl sulfate/polyacrylamide gels is slightly altered when the *groE* gene carries a missense mutation.

The second protein coded by the inserted *E. coli* DNA, the 75,000-dalton protein, is also absent in lysates of the *groE* amber phage. Unlike the 65,000-dalton protein, it is not restored by amber suppression. It is not clear why it should be affected by the *groE* amber mutation. Possibly the mutant phage carries a second, non-amber mutation that is responsible for the disappearance of the 75,000-dalton protein. Alternatively, production or stability of the 75,000-dalton protein might depend on the presence of high levels of the *groE* protein. We have been unable to distinguish between the possibilities of one or two mutations by studying revertants of the amber mutation, because apparent revertants, which arise with a frequency of about 10^{-3}, are probably the result of recombination with the host chromosome and are not true revertants.

Earlier genetic studies indicated that the ratio of gpE to *groE* protein (gp groE) may be critical to correct head assembly (3). In certain circumstances (*groE*A strains at 37°) the only λes that could be obtained to overcome the *groE* defect were mutants that reduced the amount of gpE—i.e., incompletely suppressed *E* amber mutations. It was argued that the *groE* mutation reduced the functional level of *groE* protein, and that correct assembly was restored by reducing the level of gpE and restoring the proper gpE/gp groE ratio. Whether or not this explanation is correct, the results presented here argue that changing the gpE/gp groE ratio in the opposite direction, that is, increasing the level of gp groE, is *not* detrimental to head assembly. The rate of gp groE synthesis in λgt-Ec-groE-infected *gro*$^+$ cells is severalfold higher than the rate in uninfected cells or in cells infected by wild-type phage. [This is true in cells that have not been UV-irradiated as well as in the irradiated cells shown here (our unpublished data).] Because this high rate of synthesis continues for an appreciable fraction of a cell doubling time, we conclude that the concentration of gp groE must be significantly higher in a λgt-Ec-groE infection than in a wild-type infection. Nonetheless, growth of λgt-Ec-groE on *gro*$^+$ cells is normal.

Now that the *groE* protein can be recognized, it should be possible to ask detailed biochemical questions about how it interacts with phage proteins and with other host components during phage assembly. Further, it may be possible to study what role the *groE* protein plays in the uninfected cell.

We are indebted to John Cameron and Ron Davis for making available their pool of transducing phages, Ron Davis for reading the manuscript, and Costa Georgopoulos and Barbara Hohn for communicating their results prior to publication. We thank Lorraine Lasher for expert technical assistance. This research was supported by National Institutes of Health Research Grant AI-12227 and a National Institutes of Health Research Career Development Award to R.W.H.

1. Casjens, S. R. & King, J. (1975) *Annu. Rev. Biochem.* **35**, 555–611.
2. Hershko, A. & Fry, M. (1975) *Annu. Rev. Biochem.* **35**, 775–797.
3. Georgopoulos, C., Hendrix, R., Casjens, S. & Kaiser, A. D. (1973) *J. Mol. Biol.* **76**, 45–60.
4. Sternberg, N. (1973) *J. Mol. Biol.* **76**, 25–44.
5. Georgopoulos, C., Hendrix, R., Kaiser, A. D. & Wood, W. (1972) *Nature New Biol.* **239**, 38–41.
6. Takano, T. & Kakefuda, T. (1972) *Nature New Biol.* **239**, 34–37.
7. Coppo, A., Manzi, A., Pulitzer, J. F. & Takahashi, H. (1973) *J. Mol. Biol.* **76**, 61–87.
8. Takahashi, H., Coppo, A., Manzi, A., Martire, G. & Pulitzer, J. F. (1975) *J. Mol. Biol.* **96**, 563–578.
9. Hendrix, R. (1971) in *The Bacteriophage Lambda*, ed. Hershey, A. D. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 355–370.
10. Thomas, M., Cameron, J. & Davis, R. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 4579–4583.
11. Davidson, N. & Szybalski, W. (1971) in *The Bacteriophage Lambda*, ed. Hershey, A. D., (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 45–82.
12. Cameron, J. P., Panasenko, S. M., Lehman, I. R. & Davis, R. W. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 3416–3420.
13. Georgopoulos, C. P. & Eisen, H. (1974) *J. Supramol. Struct.* **2**, 349–359.
14. Thomas, M. & Davis, R. (1975) *J. Mol. Biol.* **91**, 315–328.
15. Murialdo, H. & Siminovitch, L. (1972) *Virology* **48**, 785–823.
16. Georgopoulos, C. P. & Hohn, B. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 131–135.